Efficient CNN-based Acute Lymphoblastic Leukaemia Detection using Selective Data Augmentation and Transfer Learning

Manas Ranjan Mishra[1,*], Chinmaya Mohapatra[1], Pramod Kumar Meher[1,*], Ashish Behera[1], Akash Patel[1], Nishant Kumar Mishra[1]

C.V Raman Global University, Department of Computer Science and Engineering, Mahura, Janla, Bhubaneswar, Odisha, India, 752054

Abstract

Acute Lymphoblastic Leukemia (ALL) is a hematologic cancer characterized by a high concentration of blast cells, or immature lymphocytes. Conventionally, leukemia is diagnosed by looking under a microscope at the peripheral blood and bone marrow smear and utilizing sophisticated laboratory testing. This is a laborious and error-prone method. Therefore, automated leukemia diagnosis is required. Unlike standard deep learning approaches, transfer learning performs better in tiny databases, which is why it is becoming a popular method for analyzing medical images. In this paper, we have proposed a VGG-16 deep feature extraction-based ALL detection model followed by a custom Convolution Neural Network(CNN) model to classify the stages (Benign , Early, Pre, Pro) of Acute Lymphoblastic Leukemia (ALL) precisely. Our study centers on the identification and classification of distinct developmental stages, namely Benign, Early, Pre, or Pro, as opposed to solely discerning the malignancy status of the cells. The proposed model is trained and validated (with 80% and 20%) on the publicly available ALL image 2021 dataset. It has an classification accuracy of 99.66% and it achieves the best performance compared to the recent studies on the ALL image dataset (2021).

Keywords: Acute Lymphoblastic Leukemia, CNN, Deep learning, Transfer Learning

1. INTRODUCTION

Blood is a fluid that circulates through the circulatory system, transporting nutrients, oxygen, hormones, and waste products to and from cells. It is a vital component of the body. It is mainly composed of red blood cells (RBC), white blood cells (WBC), and platelets suspended in a liquid plasma matrix. It plays a crucial role in maintaining homeostasis, supporting the immune system, and ensuring proper oxygenation of tissues and organs[1][2]. White blood cells which make up about 1% of blood come in five different varieties, namely, monocytes, basophils, eosinophils, lymphocytes, and neutrophils contributing 60%, 30%, 5%, 4%, and less than 1% of the distribution, respectively. WBCs are essential for the immune system of the body and variations in WBC counts may signal a possible cancer risk and immune system issues. The key indicators of leukemia include an increase in blasts, or immature blood cells, and a genetic change in the bone marrow's primary cells, where blood cells are formed. This has a major effect on the blood that flows through our veins [3]. The organs, such as the bone marrow, lymph nodes, spleen, liver, brain, and central nervous system, are all affected by quickly multiplying blasts. Based on the sub-types of hematopoietic stem cells from which leukemia develops, leukemia is classified into acute lymphoblastic leukemia (ALL) and acute myeloblastic leukemia (AML) [4] [5].

The timely and precise detection of ALL is crucial for the effective treatment of patients. Traditional methods for ALL detection involve the manual examination of blood smears by trained hematologists, which can be time-consuming and prone to human-factor variability. In recent years, machine learning (ML) and deep learning (DL) techniques have been tried for the automated detection of ALL from blood cell images [6][7]. These technologies hold significant potential to improve the speed and reliability of diagnoses. Convolutional neural network (CNN) variants have emerged as powerful tools for automating medical image analysis with quicker and more accurate disease diagnoses. Transfer learning approaches based on pre-trained models are used to extract meaningful features from medical images to reduce the training complexity and enhance diagnostic efficiency. In this paper, we focus on applying transfer learning for the detection of ALL, utilizing the well-established VGG-16 architecture as a pre-trained model [8], alongside a custom CNN model [9]. The utilization of DL models in medical image analysis, particularly in the context of detecting ALL from peripheral blood smear (PBS) images, presents a promising avenue for early and accurate diagnosis[10]. The success of such models hinges not only on the quality and quantity of the available dataset but also on the ability of the model to generalize to diverse and unforeseen variations in the data. In this paper, we underscore the need and im-

*Corresponding author

Email addresses: manascvrce@gmail.com (Manas Ranjan Mishra), babunchandapur@gmail.com (Chinmaya Mohapatra), pkmeher@gmail.com (Pramod Kumar Meher), ashishbehera8382@gmail.com (Ashish Behera), akash1491625@gmail.com (Akash Patel), nishantmishra465@gmail.com (Nishant Kumar Mishra)

portance of data augmentation in ALL detection using diverse PBS images from various clinical sources. To enhance the accuracy and robustness of ALL detection, we have selected data augmentation techniques suitable for this application.

Inherent variations in staining intensity, cell morphology, and overall image characteristics necessitate exposing the model to a broader spectrum of these differences. The augmentation strategies discussed include random rotations for capturing cell orientation variability, flips (horizontal and vertical) to simulate symmetries and address potential asymmetry in cell distribution, random zooming, and scaling to replicate variations in magnification levels, and adjustments in brightness and contrast to accommodate staining intensity differences crucial in PBS images. These strategies aim to enhance the model's ability to learn invariant features and improve generalizability. The subsequent sections of the report promise to delve into the methodology, dataset, experimental results, and implications, with the ultimate goal of providing a comprehensive understanding of transfer learning's role in ALL detection and its potential to revolutionize leukemia diagnosis and treatment.

The rest of the paper is organized as follows. Section 2 highlights the key developments in ALL detection and classification techniques proposed in earlier research. Section 3 describes the ALL dataset and the data augmentation techniques used in the proposed model. In Section 4 we discuss the details of the proposed DL architecture for ALL classification. Section 5 presents the results and comparative analysis of the accuracy performance of the proposed model using confusion matrices and ROC curves to provide a comprehensive understanding of the results. Section 6 summarizes the key findings, implications for future research, and potential improvements for the model.

2. RELATED WORK

Table I provides a comparative study on ALL detection and classification, detailing the methods used, the datasets employed, and the corresponding performance scores.

Most of the work reported in the literature for classifying normal versus leukemic cells employs DL-based methods using publicly available ALL image database (ALL-IDB) [11]. CNN is the most prominent DL approach used for such classifications to address the lymphocyte blast detection and identification tasks [12][13][14]. A 16-layer CNN model was introduced by Chand *et al.* [15] for the diagnosis of ALL. They obtained an average accuracy of 98.17% in two separate experimental setups with 20 and 30 model training epochs, respectively, using the ALL-IDB1 dataset. Further, a customized CNN model was proposed by Sampathila *et al.* [16], and trained the model with the CNMC 2019 dataset. They classified leukemic blast cells and normal cells with an accuracy of 95%. Rezayi *et al.*[17] rebuilt the CNN architecture by combining ResNet-50 and VGG16 models. They obtained the ALL validation accuracy score of 84.63% in the case of CNN and reported the highest validation score of 81.72% in the case random forest (RF) algorithm using the CodaLab competition dataset.

Basymeleh et al. [18] have proposed comparative deep learning architectures from the HSV and its augmented images, for 4-class ALL classification and its diagnosis by employing the Adam optimizer. They analyzed the VGG16 architecture better than other compared deep learning models from experimental results and reported 97.50 % accuracy, 99.96% sensitivity, 100% specificity, and 98.44% validation accuracy. An advanced MobileNetV2 deep neural network proposed by Tusar *et al.* [19] to detect ALL blast cells from microscopic blood smear images. They have achieved 97 % accuracy with substantial-high specificity and sensitivity in identifying multiple ALL subtypes. Furthermore, they have implemented telediagnosis software to help physicians diagnose patients in real-time and management of this in life-threatening conditions. Gokulkrishnan *et al.* have distinguished between benign cells and the three phases of malignant ALL lymphoblast cells from four-class open source ALL dataset [20]. They have obtained 98 % ALL detection accuracy by hyperparameter tweaking on pre-trained CNNs, namely ResNet-50 and ResNet-101, to distinguish between benign cells and the three phases of malignant ALL lymphoblast cells. A hybrid model has been demonstrated by combining various ML approaches in [21]. They have suggested two distinct ALL classification models namely AlexNet hybrid and ML-based model tested using ALL-IDB2 dataset. The proposed model is outperformed by the AlexNet-based detection model, which combines feature extraction and classification, according to the results, which show 100 % classification accuracy. Saeed *et al.* have proposed an automated system for diagnosing ALL using a CNN technique, they have obtained an accuracy of 99.61% on the ALL-IDB 1 and Leukemia-IDB 2 datasets through data augmentation to mitigate overfitting [32]. Attallah *et al.* [33] have extended conventional CNN with wavelet polling followed by a Densed or LSTM layer for ALL detection. They combined deep features with handcrafted spectral features in their method and obtained 100 % ALL detection and sub-type classification.

3. Dataset and Data-Augmentation

3.1. Dataset

The publicly availabe ALL 2021 dataset consists of 3256 peripheral blood smear (PBS) images downloaded from the Kaggle website [23] .This database prepared in the bone marrow laboratory at Taleqani Hospital in Tehran, Iran from 89 patients suspected of having ALL.The dataset is divided into two categories: benign and malignant. The benign category contains hematogones, whereas the malignant category includes the ALL group with three subtypes of malignant lymphoblasts: Early Pre-B, Pre-B, and Pro-B ALL. Images were taken using a Zeiss camera mounted on a microscope at 100x magnification and saved in JPG format. A specialist utilized flow cytometry to accurately identify the types and subtypes of the cells. Additionally, it provides segmented images produced through color thresholding-based segmentation in the HSV color space.
Zeiss cameras mounted on microscopes provided excellent magnification (up to 100x) for the data frames included in the

Table 1: Related work studies on recent in Acute Lymphoblastic Leukemia Classifications

References	Database used	Data Pre-processing	Models used	Result Scores
Chand *et al.*[15]	ALL-IDB1	No preprocessing	16-layer CNN	Accuracy- 97.95%
Sampathila *et al.*[16]	CNMC 2019	Color thresholding, augmentation (flipping, rotation, brightness, blur, noise)	Custom CNN (ALLNET)	Accuracy- 95%
Rezayi *et al.*[17]	CodaLab competition dataset	Normalization, segmentation	ResNet50, VGG16, Novel CNN	Validation accuracies- 84.625% (VGG16)
Basymeleh *et al.*[18]	ALL image Kaggle	Resizing, augmentations (rescaling, rotation, shift, shear, zoom, flip)	Various deep learning models	Accuracy- 97.50% (VGG16)
Tusar *et al.*[19]	ALL image Kaggle	Resizing,padding,augmentation, normalization	Various deep neural networks	Accuracy: 97%(MobileNetV2)
Gokulkrishnan *et al.*[20]	ALL image Kaggle	Color thresholding for segmentation	ResNet-50 and ResNet-101	Validation Accuracy: 98.62% (ResNet-101)
Magpantay *et al.*	ALL IDB-1 and ALL IDB-2	Annotation of images	YOLOv3	Training Accuracy : 97.2% and mAP : 99.8%
Shashank *et al.*[22]	ALL-IDB2	Three-fold augmentation (rotation, zooming, flipping), resizing, normalization	Hybrid model: AlexNet + ML classifiers	Accuracy- 100%
Saeed *et al.*[32]	ALL-IDB1 & IDB2	Data augmentation	CNN	Accuracy- 99.61%
Omneya *et al.*[33]	ALL-IDB2 & ALL image Kaggle	Handcraft and Deep Hybrid Features	CNN-LSTM	Accuracy- 100%

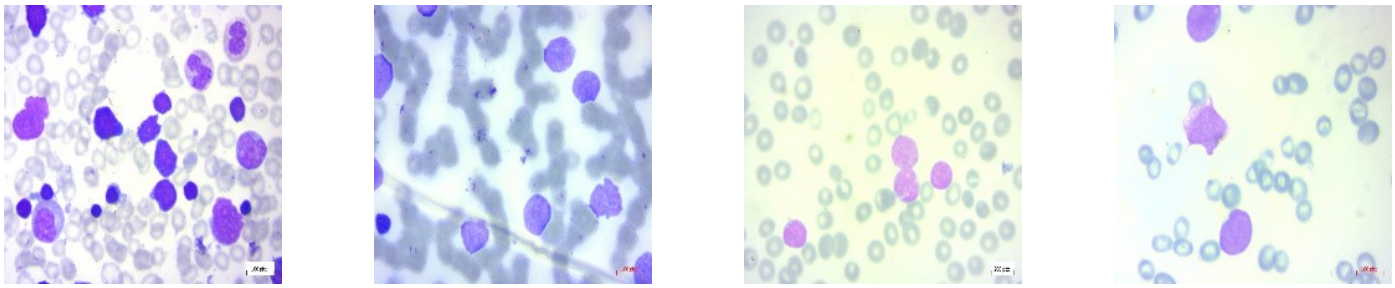
dataset. To be easily accessible for research needs, these photographs are stored in JPG format. To further improve its applicability for research experimentation, the dataset was preprocessed. Specifically, segmentation in the HSV color space was done using color thresholding as part of this preprocessing, and segmented images were produced. These segmented images are essential for the creation and evaluation of machine learning and image processing algorithms that detect and classify ALL cells automatically.

Table 2: Input matrix of ALL Image dataset [23] for classification

# input images	**benign**	**early**	**pre**	**pro**
training (t)	2021	3950	3862	3223
validation (v)	204	216	240	216
total (t+v)	2225	4166	4102	3439

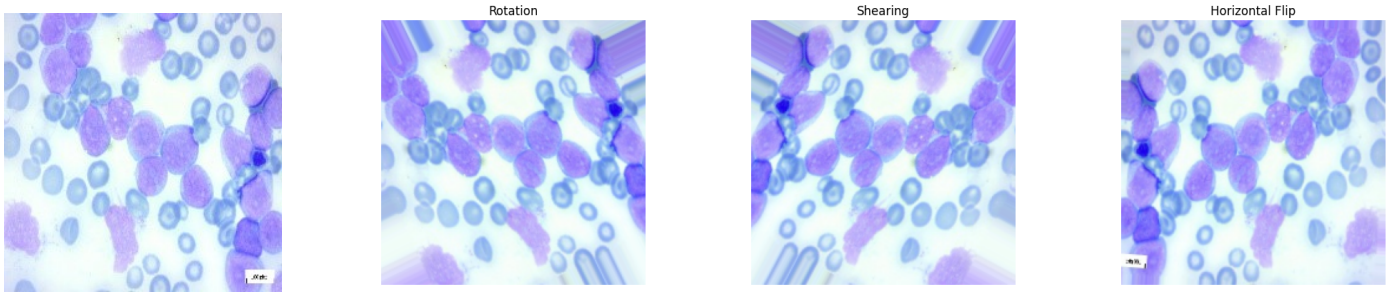
3.2. Data Augmentation

In this section we have highlighted the benefits of using multiple methods of data augmentation types is a major concern in ML field to address the challenges class imbalance and insufficient training data [24]. From meticulous studies, it has been shown that the availability of massive duplicate data produced by data augmentation techniques might occasionally cause the medical picture diagnostic models to become overfit. One typical job in machine learning and computer vision is data augmentation . The use of vision-based AI models in biomedical illness detection relies heavily on the datasets that are made accessible. These models incorporate a variety of medical imaging data, including X-ray, MRI, CT scan, histopathology, and ultrasound pictures. In medical imaging, data augmentation aims to provide additional samples, which are precisely distinct cases from the same hospital using the same imaging device. Even though there are several DL techniques available, data augmentation is utilized to increase accuracy and predict or identify illnesses. The common issue with these models is that they do not always produce good results. This might be due to inadequate data or incorrect procedural data-augmentation actions being taken. Augmenting data is necessary for machine learning for several significant reasons. First off, by artificially growing the dataset, assists in getting around restrictions on the availability of data. This is particularly important in fields where obtaining vast quantities of labeled data is costly or challenging. Second, it enhances performance on unknown samples and lessens overfitting by exposing the model to a greater range of data, which promotes generalization. Thirdly, augmentation may be used to improve model accuracy and fairness in problems when the class distributions are unbalanced. It does this by creating artificial instances of minority classes. Furthermore, in medical imaging like leukemia diagnosis, augmentation helps in establishing a robust training model by virtue of its sample



(a) (b) (c) (d)

Figure 1: Different Stages of ALL : (a) Benign, (b) Early, (c) Pre, (d) Pro



(a) (b) (c) (d)

Figure 2: (a) Original Image, (b) After Rotation , (c) After Shearing, (d) Horizontal Flip.

variation properties. Data augmentations can be considered to produce better results in time-critical applications like disease diagnosis and predictions.

Using image duplication as a data augmentation technique for blood cell images may lead to faster training but risks overfitting, where the model memorizes specific instances rather than learning general patterns. Additionally, it may not effectively capture the true variability in blood cell images, limiting the model's ability to generalize to unseen data. There's also a concern that duplication could amplify biases in the dataset, potentially skewing predictions. Therefore, while image duplication can quickly increase dataset size, it should be used cautiously alongside other augmentation methods to ensure robust and unbiased model performance.

In our research work, we have used the following techniques for the generation of data to be augmented. Horizontal and vertical flips simulate variations in left-right and top-bottom symmetries, respectively. The images in the dataset undergo augmentation with a shearing parameter of 0.2 and a zoom range parameter of 0.2, contributing to a mild shearing effect and variability in zoom levels. The rescaling parameter is set to 1./255, normalizing pixel values to a range between 0 and 1, facilitating effective model training and convergence. These methods introduce diverse variations into the training dataset, preventing overfitting and enhancing the model's ability to generalize to unseen data. They simulate real-world scenarios such as different orientations, lighting conditions, and staining techniques, ensuring the model's robustness. Additionally, they make the model more resilient to noise and distortions common in microscopy imaging, improving its performance in medical diagnostic settings. Integrating these techniques enables the development of more versatile and reliable models, advancing medical diagnostics and hematology research.

The sequence of variation of images to generate the augmented dataset is depicted in Fig 2. Table 2 shows the input matrix of the lymphoblastic leukemia dataset for classification showing the number of images in the dataset after augmentation. The dataset is partitioned into training and validation sets. The training set contains 2021 benign images, 3950 early-stage cancer images, 3862 pre-cancerous images, and 3223 precancerous images. The validation set contains 204 benign images, 216 early-stage cancer images, 240 pre-cancerous images, and 216 precancerous images. Figure 3 visualizes the data augmentation process used in proposed ALL detection model.

4. Proposed Model

4.1. VGG-16

We discussed here the adaptation of the VGG-16 model for our image classification task [27]. The VGG-16 model was initially designed for the ImageNet Large Scale Visual Recognition Challenge. It is popularly used for its simplicity, using small 3x3 convolutional filters and max-pooling layers to effectively capture hierarchical features in images.

The proposed VGG-16 model is constructed with a total of 16 layers, organized into five convolutional blocks, each followed by a max-pooling layer. The layers in each block have progressively increasing filter depth, allowing the model to capture both low-level and high-level features. The convolutional blocks are as follows:

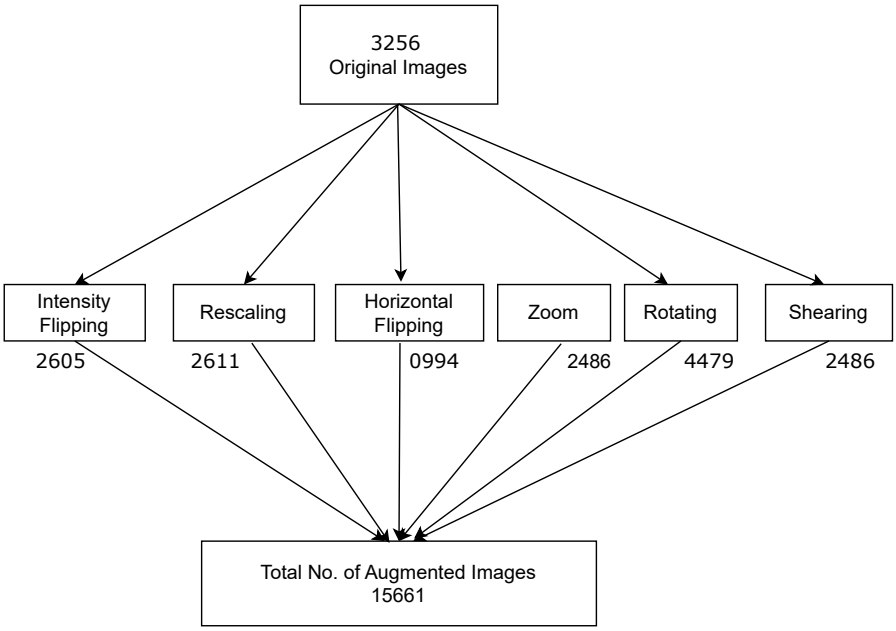


Figure 3: Flowchart of the model execution

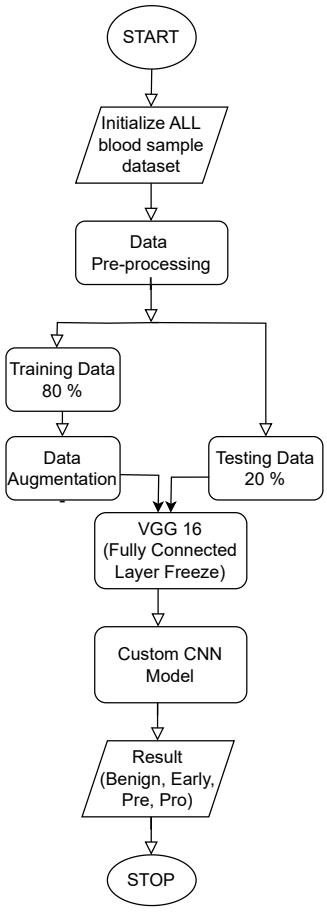


Figure 4: Flowchart of Execution Model

1. The first convolutional block consists of two 3x3 convolutional layers, each followed by a max-pooling layer. These layers are designed to capture basic features like edges and textures.
2. The second and third convolutional blocks follow a similar pattern, with the depth of filters increasing, enabling the detection of more complex features.
3. The fourth convolutional block is composed of three convolutional layers, which further increase the depth of feature maps. This helps in identifying even more intricate patterns in the input images.
4. The fifth and final convolutional block contains three convolutional layers, and it plays a critical role in extracting high-level features.

To adapt the VGG-16 model for our specific task, we fine-tune it by freezing the pre-trained weights. This ensures that the model retains the knowledge acquired during its training on the ImageNet dataset. We then add 16 custom layers in our experiment, on the top of the VGG-16 base to enable it to classify images according to our specific target classes. These custom layers include Flatten layers for converting the 2D feature maps into 1D vectors, Dense layers for fully connected layers, Batch-Normalization for regularization, and Dropout layers to reduce overfitting. These custom layers enable the VGG-16 model to be fine-tuned to our image classification task effectively. Refer to Figure 4 for a visual representation of the program flowchart, illustrating how data flows through the combined architecture.

4.2. CNN

CNNs are sophisticated multi-layered architectures capable of autonomously extracting intricate, high-dimensional features from extensive datasets [25][26]. By leveraging the interactions among their fundamental components, neurons, they ana-

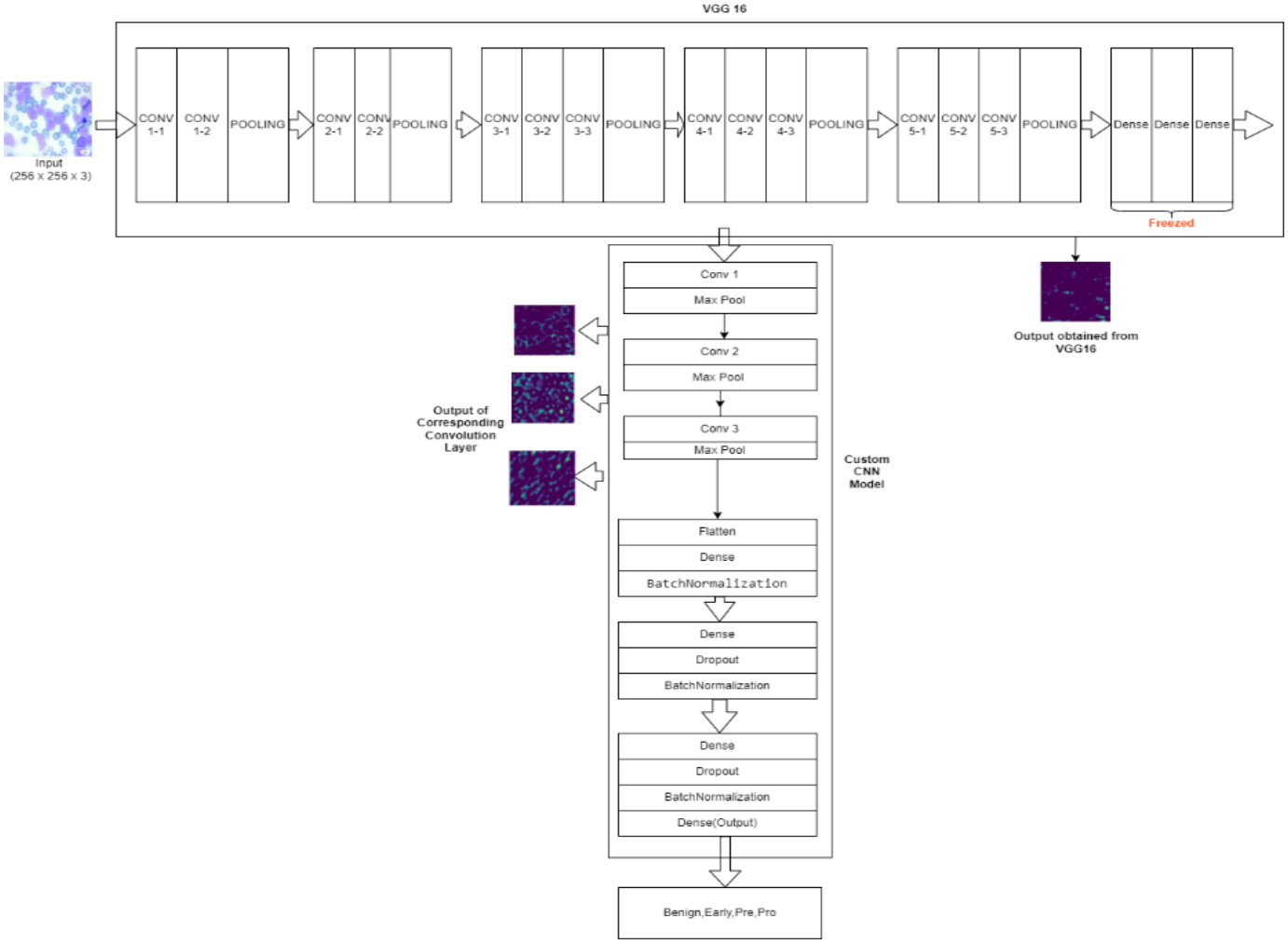


Figure 5: Architecture of the proposed model

lyze images and acquire patterns, mirroring the learning mechanisms of the human brain. CNNs possess the capacity to learn and extract relevant characteristics for classification automatically. Moreover, unlike earlier machine learning methodologies, automatic blood cell classification through CNNs does not explicitly depend on complex cell region segmentation and subsequent feature selection.

The initial segment of a CNN always consists of the input layer, responsible for processing the pixels of the image. Following this, convolutional layers are employed, comprising filters designed to recognize specific patterns within images and extract features, resulting in sets of feature maps. Convolutional blocks, integral to CNNs, consist of one or more convolutional layers depending on the architecture. Subsequently, a pooling layer typically follows a convolutional layer or block, reducing the dimensionality of feature maps while preserving crucial information and eliminating extraneous features. The widely used max pooling function minimizes the size of feature maps by selecting maximum values. This combination of convolutional and pooling layers is iteratively applied throughout the network architecture to progressively extract more complex information[22] . The depth of a network is determined by the repetition of these layers, yielding a collection of feature maps reflective of the input images.

In this study, we employed the fine-tuning approach, which involves adapting a pre-existing CNN architecture trained for a different purpose. By utilizing ALL image 2021 dataset [23], we modified and re-trained certain layers to create a comprehensive classifier . A schematic illustrating the interconnected levels of the integrated model is depicted in Figure 5. The objective was to address the scarcity of training instances by leveraging the information from a CNN pre-trained with a substantial dataset. In our investigation, this pre-training also served as an effective weight initialization for subsequent image classification tasks.

4.3. Proposed Transfer Learning Model

In our pursuit of effective ALL detection from Peripheral Blood Smear (PBS) images, we propose a sophisticated algorithm based on transfer learning with the VGG16 convolutional neural network architecture. The algorithm commences with the initialization of the VGG16 base model, omitting its top layers and configuring the input shape to (IMG SIZE, IMG SIZE,

3). The decision to exclude the top layers aligns with our objective of fine-tuning the model for the specific complexities of ALL detection, leveraging the pre-existing knowledge captured in the lower layers of VGG16.

The weights of the base model are then frozen, ensuring that the foundational features learned by VGG16 remain intact during subsequent training. Our custom layers, designed to augment the capabilities of the base model, are meticulously constructed. Three convolutional layers are introduced, each with 3x3 filters and ReLU activation, employing 'same' padding to retain spatial information. Max-pooling layers follow each convolutional operation, aiding in down-sampling and feature selection. This architectural choice is deliberate, aiming to capture hierarchical features inherent in PBS images.

Subsequent to the convolutional layers, the algorithm incorporates flattening before introducing three dense layers. The first two dense layers consist of 512 units with ReLU activation, accompanied by batch normalization for improved convergence and dropout for regularization (0.1 and 0.2, respectively). The final dense layer, serving as the output layer, aligns with the categorical nature of the diagnostic task, employing softmax activation for multi-class classification.

In totality, the model comprises the VGG16 base model and eight additional layers, contributing to a total of eleven layers in the custom architecture. This amalgamation is crucial for striking a balance between leveraging pre-existing knowledge and allowing the model to adapt to the nuances of ALL detection.

For training and validation, the algorithm is compiled using the Adam optimizer, categorical crossentropy loss, and accuracy as the evaluation metric. Training is conducted over 25 epochs and validation is performed on a dedicated dataset to monitor convergence and prevent overfitting. The validation phase entails assessing the model's accuracy on an independent test dataset, reflecting the algorithm's proficiency in discerning features indicative of ALL within PBS images.

By combining these two models, we aim to improve the overall performance of our image classification system and achieve higher accuracy in classifying images into our target categories. The 'combined model' architecture enables feature extraction at multiple scales and levels of abstraction, enhancing the model's ability to discriminate between different image classes.

5. Result and Analysis

Our transfer learning model combines a pre-trained VGG-16 model with a custom CNN architecture, creating a powerful image classification framework. The VGG-16 model, with its depth and proven capabilities in feature extraction, serves as a strong foundation. By obtaining extra specialized characteristics, the custom CNN enhances the VGG-16's capabilities to better identify and categorize objects in photos. The model's performance is promising, with its ability to learn and adapt to a wide range of image data. By leveraging the pre-trained weights of VGG-16, the model benefits from the knowledge gained through extensive training on a large image dataset. The additional custom CNN layers further fine-tune the model to the specific classification task, improving its accuracy and robustness. The transfer learning model was trained for a total of 25 epochs to assess its performance and convergence. Table 3 provides model's training and validation results, offering insights into its learning curve and generalization capabilities.

Table 3: Training and validation results

Epoch Size	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
25	0.9932	0.9966	0.0195	0.0112

In the context of architecture, our CNN model integrates the strengths of transfer learning with VGG16 as its base, incorporating custom convolutional layers and essential architectural elements. By leveraging VGG16's established convolutional blocks without the fully connected layers, our model capitalizes on its ability to capture intricate features from images effectively. The addition of convolutional and pooling layers with 'same' padding ensures preservation of spatial hierarchy and enables robust feature extraction at multiple scales. BatchNormalization and Dropout layers are integrated to stabilize and regularize the training process, enhancing our model's resilience against overfitting. The softmax activation in the output layer facilitates multi-class classification, aligning with the categorical nature of our dataset. This architecture is meticulously designed to balance complexity with computational efficiency, making it suitable for tasks requiring precise image classification. Moreover, it accommodates adjustments in layer configurations and augmentation strategies to adapt to diverse dataset characteristics and specific training requirements. Our model is trained over 25 epochs to optimize performance and convergence, balancing training time with the need to capture intricate features from the dataset effectively. With a base model like VGG16, which has approximately 138 million parameters, our architecture benefits from its depth and complexity in feature extraction. The inclusion of additional custom layers adds further flexibility and specificity to the model's learned representations. During training, the simulation time is carefully managed to ensure efficient use of computational resources while achieving high accuracy in image classification tasks. This approach not only maximizes learning from available data but also ensures scalability and adaptability for future model enhancements and applications.

To assess the performance of our model comprehensively, we employed a range of evaluation metrics, including:

• Confusion Matrix: The confusion matrix provides a detailed breakdown of the model's predictions, including true positives, true negatives, false positives, and false negatives. It is a valuable tool for understanding the model's ability to classify different classes correctly. Refer to Figure 6 visual representa-

Table 4: Classification Report

	Precision	Recall	F1-score
Benign	1.00	1.00	1.00
Early	1.00	1.00	1.00
Pre	1.00	0.99	0.99
Pro	0.99	1.00	1.00
Accuracy			1.00

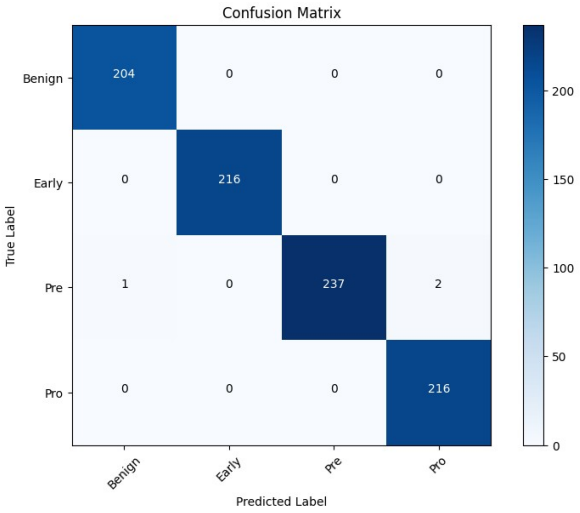


Figure 6: Classwise Visualization of confusion matrix

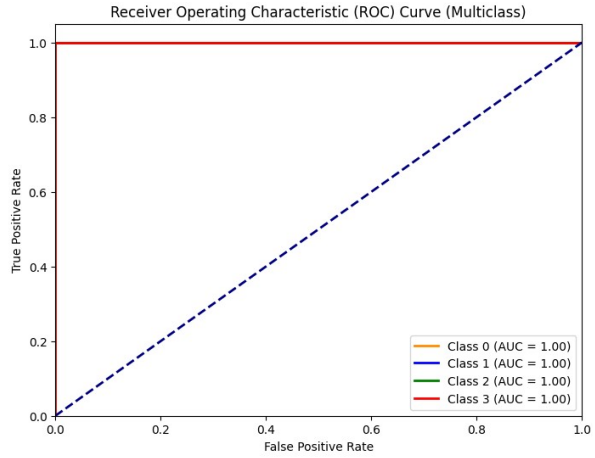


Figure 7: ROC Curve

tion of the confusion matrix , Figure 7 for ROC Curve, Classification Report (Table 4), Training and validation accuracy per epoch(Figure 8).

• F1 Score: The F1 score is a balance between precision and recall, offering a single metric to evaluate the model's overall performance in terms of both false positives and false negatives.

F1 score = 2 * (precision * recall)/ (precision + recall)

• Accuracy: Accuracy measures the overall correctness of the model's predictions. It provides an indication of how well the model performs across all classes.

Table 5: Example table with five columns

Reference	Dataset used	Pre-processing	Model used	Result Achieved
[18]	Publicly available dataset	Resizing, augmentation	Various deep learning models	Accuracy: 97.50%, (VGG16)
[19]	Publicly available dataset	Resizing, padding, augmentation, normalization	Various deep neural networks	Accuracy: 97.42% (MobileNetV2)
[20]	Publicly available dataset	Color thresholding for segmentation	ResNet-50 and ResNet-101	Validation Accuracy: 98.62% (ResNet-101)
Proposed Model	Publicly available dataset	Rotation, shearing, horizontal flipping	VGG-16 and CNN Fusion Transfer Learning Model	Accuracy : 99.66%

Accuracy = (TP + TN) / (TP + FP + FN + TN)

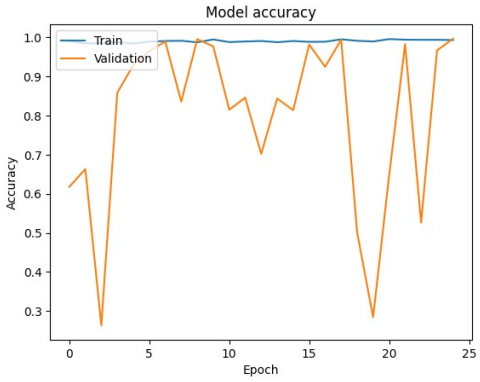
• Precision: Precision is a metric that evaluates the correctness of positive predictions made by the model. It is useful in scenarios where false positives need to be minimized.

Precision = TP/ TP+FP

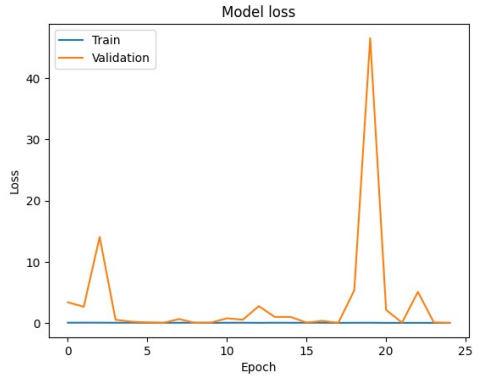
• Sensitivity (Recall): Sensitivity, also known as recall, measures the model's capability to correctly identify true positives. It is vital for situations where false negatives are highly undesirable.

Sensitivity = TP / FN+TP

The validation accuracy of the model is reported as 99.66%, highlighting its proficiency in making correct predictions across all classes. These results collectively underscore the robust performance of the model in classification tasks, with particularly noteworthy precision and recall values across diverse categories. Refer to Table 5 for comparison of evaluation parameters of related work on detection of acute lymphoblastic leukemia and the proposed model.



(a)



(b)

Figure 8: Training and validation (a) accuracy (b) loss curve

6. Conclusion

By using CNN and transfer learning techniques, this study offers a reliable methodology for the automated detection and classification of ALL. Blending the VGG-16 architecture with a customized CNN, our model demonstrates better performance in differentiating between Benign, Early, Pre, and Pro stages of ALL compared to the existing learning models. Impressively, the proposed model achieves an accuracy rate of 99.66%, which is significantly higher that state of the art machine learning models.

By harnessing transfer learning, we leveraged the pre-trained VGG-16 model's feature extraction capabilities and further fine-tuned it for our specific classification task. Data augmentation techniques were employed to mitigate data scarcity issues and improve generalization. The comprehensive evaluation metrics, including the confusion matrix, F1 score, accuracy, precision, and sensitivity, highlight the model's effectiveness. This research contributes significantly to the field of medical image analysis and diagnostics, with the potential to expedite the diagnosis of ALL, reduce human error, and improve patient outcomes.

Future work may involve exploring advanced data augmentation techniques and experimenting with other deep learning architectures. Our ultimate goal is to provide healthcare professionals with a powerful tool for early and accurate ALL detection, thereby advancing patient care and contributing to the battle against leukemia.

References

[1] Acute Lymphocytic Leukemia. Available from: ncbi.nlm.nih.gov. https://www.ncbi.nlm.nih.gov/books/NBK459149/, [Accessed 01-11-2023].

[2] F. J. Schiffman, Hematologic pathophysiology. Lippincott Williams & Wilkins, 1998.

[3] V. P. Burke and J. M. Startzell, "The leukemias," Oral and Maxillofacial Surgery Clinics of North America, vol. 20, no. 4, pp. 597-608, 2008.

[4] Vaghela HP, Modi H, Pandya M, Potdar MB (2015) Leukaemia detection using digital image processing techniques. Leukaemia 10(1):43–51.

[5] Chatap N, Shibu S (2014) Analysis of blood samples for counting leukaemia cells using support vector machine and nearest neighbour. IOSR J Comput Eng (IOSR-JCE) 16(5):79–87.

[6] Mahmood, Nasir, et al. "Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach." Medical Biological Engineering Computing 58 (2020): 2631-2640.

[7] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimedia Tools and Applications, 82(17), 26731–26769.

[8] Yu, X., Wang, J., Hong, Q.Q., Teku, R., Wang, S.H., & Zhang, Y.D. (2022). Transfer learning for medical images analyses: A survey. Neurocomputing, 489, 230–254.

[9] Raina, R., Gondhi, N., Chaahat, Singh, D., Kaur, M., & Lee, H.N. (2023). A Systematic Review on Acute Leukemia Detection Using Deep Learning Techniques. Archives of Computational Methods in Engineering, 30(1), 251–270.

[10] Miko.lajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117–122).

[11] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, & Azamossadat Hosseini. (2021). Acute Lymphoblastic Leukemia (ALL) image dataset.

[12] Putzu, Lorenzo, Giovanni Caocci, and Cecilia Di Ruberto. "Leucocyte classification for leukaemia detection using image processing techniques." Artificial intelligence in medicine 62.3 (2014): 179-191.

[13] Rehman, Amjad, et al. "Classification of acute lymphoblastic leukemia using deep learning." Microscopy Research and Technique 81.11 (2018): 1310-1317.

[14] Saeed, Umair, et al. "DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification." Multimedia Tools and Applications 83.7 (2024): 21019-21043.

[15] Chand, S., & Vishwakarma, V. P. (2022). A novel deep learning framework (DLF) for classification of acute lymphoblastic leukemia. Multimedia Tools and Applications, 81(26), 37243-37262.

[16] Sampathila, N., Chadaga, K., Goswami, N., Chadaga, R. P., Pandya, M., Prabhu, S., ... & Upadya, S. P. (2022, September). Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images. In Healthcare (Vol. 10, No. 10,p. 1812). MDPI.

[17] Rezayi, S., Mohammadzadeh, N., Bouraghi, H., Saeedi, S., & Mohammadpour, A. (2021). Timely diagnosis of acute lymphoblastic leukemia using artificial intelligence-oriented deep learning methods. Computational Intelligence and Neuroscience, 2021.

[18] Basymeleh, Aiman Muhamad, Bagus Esa Pramudya, and Reinato Teguh Santoso. "Acute Lymphoblastic Leukemia Image Classification Performance with Transfer Learning Using CNN Architecture." 2022 4th International Conference on Biomedical Engineering (IBIOMED). IEEE, 2022.

[19] Tusar, MTH Khan, et al. "An Intelligent Telediagnosis of Acute Lymphoblastic Leukemia using Histopathological Deep Learning." J. Comput. Theor. Appl 2.1 (2024): 1-12.

[20] Gokulkrishnan, Nitla, Tushar Nayak, and Niranjana Sampathila. "Deep Learning-Based Analysis of Blood Smear Images for Detection of Acute Lymphoblastic Leukemia." 2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE, 2023.

[21] Magpantay, Leo Dominick C., et al. "A transfer learning-based deep CNN approach for classification and diagnosis of acute lymphocytic leukemia

cells." 2022 International conference on decision aid sciences and applications (DASA). IEEE, 2022.
[22] Surya Sashank, G. V., Jain, C., & Venkateswaran, N. (2021). Detection of acute lymphoblastic leukemia by utilizing deep learning methods. In Machine Vision and Augmented Intelligence—Theory and Applications: Select Proceedings of MAI 2021 (pp. 453-467). Springer Singapore.
[23] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, and Azamossadat Hosseini. (2021). Acute Lymphoblastic Leukemia (ALL) image dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/2175623
[24] Mikołajczyk, Agnieszka, and Michał Grochowski. "Data augmentation for improving deep learning in image classification problem." 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018.
[25] Liu, Weibo, et al. "A survey of deep neural network architectures and their applications." Neurocomputing 234 (2017): 11-26.
[26] Berezsky, Oleh, et al. "Synthesis of Convolutional Neural Network architectures for biomedical image classification." Biomedical Signal Processing and Control 95 (2024): 106325.
[27] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
[28] G Litjens , T Kooi , BE Bejnordi , AAA Setio , F Ciompi , M Ghafoorian , et al. , A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88 .
[29] Sornam, M., Kavitha Muthusubash, and V. Vanitha. "A survey on image classification and activity recognition using deep convolutional neural network architecture." 2017 ninth international conference on advanced computing (ICoAC). IEEE, 2017.
[30] J Yosinski , J Clune , Y Bengio , H. Lipson , How transferable are features in deep neural networks? in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328 .
[31] A. Geron , Hands-on Machine Learning with Scikit-Learn, Keras, and Tensor- Flow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, 2019 .
[32] Saeed, Umair, et al. "DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification." Multimedia Tools and Applications 83.7 (2024): 21019-21043.
[33] Attallah, Omneya. "Acute lymphocytic leukemia detection and subtype classification via extended wavelet pooling based-CNNs and statistical-texture features." Image and Vision Computing 147 (2024): 105064.